ISBN 978-602-5823-52-7

Dr. Abdul Kodir
Nanang Ismail, MT
Asep Solih Awaluddin, M.Si
Prof. Dr. Uus Ruswandi, M.Pd

Kompresi Berbagai Bahasa Lokal Indonesia Berbasis Teori Informasi & Coding

PUSAT PENELITIAN DAN PENERBITAN
UIN SGD BANDUNG

KOMPRESI BERBAGAI BAHASA LOKAL INDONESIA BERBASIS TEORI INFORMASI DAN *CODING*

Dr. Abdul Kodir, M. Ag
Nanang Ismail, MT
Asep Solih Awaluddin, M. Si
Prof. Dr. Uus Ruswandi, M. Pd

Pusat Penelitian dan Penerbitan
UIN SGD Bandung

KOMPRESI BERBAGAI BAHASA LOKAL INDONESIA BERBASIS TEORI INFORMASI DAN *CODING*

Penulis:
Dr. Abdul Kodir, M. Ag
Nanang Ismail, MT
Asep Solih Awaluddin, M. Si
Prof. Dr. Uus Ruswandi, M. Pd

ISBN: 978 – 602 – 5823 – 52 – 7

Penyunting:
Nanang Ismail, MT

Desain Sampul dan Tata letak:
Ahmad Sujana

Penerbit:
Pusat Penelitian Dan Penerbitan UIN SGD Bandung
Jl. H.A. Nasution No. 105 Bandung
Tlp. (022) 7800525, Fax. (022) 7800525
http://lp2m.uinsgd.ac.id

Cetakan pertama, Oktober 2018

KATA PENGANTAR

Puji syukur kepada Allah Yang Maha Esa, yang senantiasa melimpahkan nikmat dan kasih sayang kepada makhluk-Nya. Atas limpahan nikmat dan anugerah-Nya pulalah, yang mengantarkan dan membimbing peneliti menyelesaikan penelitian yang berjudul Kompresi Berbagai Bahasa Lokal Indonesia Berbasis Teori Informasi dan Coding untuk Antisipasi Tindakan Terorisme dan menuangkan dalam bentuk buku ini.

Buku ini merupakan salahsatu output penelitian yang didanai oleh DIPA-BOPTN UIN Bandung Tahun 2018 melalui LP2M.

Ucapan terima kasih Kami sampaikan kepada semua pihak yang telah dengan tulus ikhlas membantu dalam penyelesaian penelitian ini. Secara khusus Kami sampaikan terima kasih kepada:

1. Pimpinan universitas melalui LP2M UIN Sunan Gunung Djati Bandung atas kepercayaannya.
2. Dekan Fakultas Saintek UIN SGD Bandung
3. Narasumber dalam penelitian
4. Tim Pengolah data yang membantu dalam pelaksanaaan penelitian
5. Semua pihak yang telah membantu baik secara langsung maupun tidak langsung dalam penelitian ini

Dengan segala kerendahan hati kami menyadari masih banyak kekurangan dan kelemahan dalam penelitian ini. Kami menerima kritik dan saran demi perbaikan penelitian-penelitian ke depan.

Akhir kata, semoga laporan ini bermanfaat bagi pembaca semuanya.

Bandung, Oktober 2018

Peneliti

DAFTAR ISI

DAFTAR GAMBAR

DAFTAR TABEL

BAB I. Pendahuluan

1.1. Latar Belakang

Salahsatu teknologi yang saat ini menjadi lokomotif era digital adalah teknologi telekomunikasi dan kompressi data. Telekomunikasi berkembang sangat pesat, dan diantara teknologi pendukungnya adalah teknik kompresi. Pada proses telekomunikasi seluler, akan terjadi perebutan kanal komunikasi antar *end system*. BTS akan memilih prioritas pemberian kanal berdasarkan beberapa pertimbangan, diantaranya ukuran data yang kecil. Teknik kompresi memegang peranan penting untuk hal ini, sehingga data yang dikomunikasikan lebih kecil. Kompresi data berhubungan dengan pengkodean teks digital, sinyal audio atau video dengan jumlah bit minimum, sehingga jumlah bit yang ditransmisikan dapat ditingkatkan dan dibawa pada sistem komunikasi dengan kapasitas lebih rendah, menghabiskan lebih sedikit ruang penyimpanan, dan memerlukan lebih sedikit bandwidth untuk transmisi yang efisien [1]. Pada jaringan sensor [2], kompresi diperlukan untuk menghemat konsumsi energi. Di antara teknik kompresi, pengkodean Huffman adalah skema pengkodean awal yang optimal, yang menjamin *decodability* unik dari simbol yang terkompresi [3]. Kode tersebut dirumuskan oleh Huffman di awal tahun 1950an dan memberikan hasil yang optimal meski kodenya tidak universal [4].

Level kompesi dipengaruhi simbol dan jumlah simbol bahasa yang digunakan dalam data teks. Indonesia memiliki banyak sekali bahasa lokal/daerah. Diantara bahasa lokal itu adalah Bahasa Sunda, Jawa, Minang, Bali, Bugis. Khoirul Anwar dalam [3] sudah menganalisis level kompressi 3 bahasa lokal yaitu Sunda, Jawa, dan Bali, tetapi banyak simbol yang tidak muncul. Setiap bahasa memiliki simbol tambahan yang unik dibandingkan dengan simbol-simbol yang digunakan dalam Bahasa Indonesia. Dikaitkan dengan teknologi kompresi yang dibahas sebelumnya, setiap bahasa memiliki tingkat kompressi

yang berbeda karena simbol dan jumlah simbol yang digunakan juga berbeda. Ukuran sumber yang jadi bahan analisis juga berpengaruh terhadap keakuratan kompresi. Semakin banyak simbol yang muncul/tercover, semakin akurat hasil yang didapat.

Ke depan, dengan akan berkembangnya teknologi 5G, banyak device akan bisa saling "bicara" dan berkomunikasi, bukan hanya dengan bahasa internasional, bahkan menggunakan berbagai bahasa lokal. Sebagai contoh, komunikasi ini terjadi di lingkungan pariwisata yang akan mempertemukan berbagai bahasa [3]. Oleh karena itu, perlu diteliti, bahasa lokal mana yang lebih optimal kompresinya, tentunya dengan sebanyak mungkin simbol yang dilibatkan. Selain itu, antar bahasa yang ada, juga memiliki tingkat similaritas yang berbeda. Dalam komunikasi, jika dua bahasa berkomunikasi, maka tingkat error yang dihasilkan akan lebih sedikit jika digunakan 2 bahasa dengan tingkat similaritas yang tinggi. Oleh karenanya, perlu juga dilakukan penelitian mengenai tingkat similaritas bahas yang ada.

Teknik kompresi ini juga bisa dikaitkan dengan solusi keamanan dan pencegahan terorisme. Sebagaimana diketahui, bahwa salah satu isu yang belakangan muncul secara nasional adalah isu keamanan dan terorisme. Kegiatan terorisme sudah melibatkan berbagai suku dengan bahasa komunikasi yang berbeda-beda. Banyak pendekatan yang digunakan untuk mencegah terorisme, termasuk pendekatan ekonomi, budaya, agama, hukum dan teknologi.

Gambar 1. Pendekatan untuk pencegahan terorisme

Secara teknologi, terorisme dapat diantisipasi dengan mengembangkan berbagai perangkat dan teknologi untuk antisipasinya. Kompressi data, secara tidak langsung berperan pada keamanan, karena data yang dikompress hakikatnya sudah terenkripsi. Selain itu pada proses pemilihan kanal komunikasi dapat diatur dengan sistem bloking bagi pihak-pihak yang "suspect" terorisme. Komunikasi yang terkompress akan diberikan prioritas kanal, sementara yang tidak terkompress dengan metoda tertentu akan diblok/diabaikan

1.2. Perumusan Masalah dan Ruang Lingkup

Setiap bahasa memiliki simbol yang berbeda, dan simbol yang berbeda ini berpengaruh terhadap level kompressi tiap bahasa. Dengan kemungkinan akan ada komunikasi antar devais dengan berbagai bahasa lokal dan internasional yang ada pada masa yang akan datang, khsusunya ketika teknologi 5G digelar, maka perlu dilakukan penelitian mengenai level kompressi dari tiap bahasa lokal dan bahasa internasional. Sehingga bisa diketahui kompressi berbagai

bahasa lokal Indonesia berdasarkan teori informasi dan *coding*, yang pada akhirnya dapat diketahui bahasa lokal mana yang lebih optimal kompressinya untuk memenuhi proses komunikasi dengan *low power consumption.* Untuk memperoleh hasil kompresi yang optimal, analisis bahasa dilakukan menggunakan *source* naskah yang memiliki ukuran lebih besar di Banding yang sudah dilakukan K. Anwar pada [3], sehingga meng*cover* jumlah simbol yang lebih banyak.

Selain itu, error dalam komunikasi antar dua bahasa bisa ditekan jika keduanya memiliki tingkat similaritas tinggi. Oleh karenanya, perlu dilihat tingkat similaritas antar bahasa lokal Indonesia. Pendekatan statistik dapat digunakan untuk mencari tingkat similaritas antar bahasa tersebut.

Selanjutnya, kompresi bahasa ini dapat digunakan untuk mengantisipasi tindakan terorisme dengan meminimalisir komunikasi antar pihak-pihak yang *suspect* terorisme. Proses pemilihan kanal komunikasi dapat diatur dengan sistem bloking bagi pihak-pihak yang *suspect* terorisme. Komunikasi yang terkompress akan diberikan prioritas kanal, sementara yang tidak terkompress akan diblok/diabaikan.

Agar tidak melebar, ruang lingkup penelitian ini dibatasi sebagai berikut:

- Metode kompressi yang akan digunakan adalah *Huffman Code dan Run Length-Huffmun code.*
- Bahasa yang akan dianalisis adalah: bahasa dan aksara Sunda, Bahasa dan aksara Jawa,.
- Analisis tingkat kompresi menggunakan *marginal probability*, membandingkan tingkat kompressi dalam percakapan menggunakan *outage probability.*

1.3. Tujuan Penelitian

Penelitian ini bertujuan melakukan analisis tingkat kompressi berbagai bahasa lokal berdasarkan teori informasi dan menentukan tingkat kompressi paling optimal dari berbagai bahasa lokal di Indonesia dengan pendekatan statistik *outage probability.*. Standar kompresi yang diperoleh selanjutnya akan disimulasikan dengan software untuk menghasilkan prototipe Alpha (skala lab) dari sistem.

1.4. Urgensi dan Kontribusi

- Kontribusi penelitian ini diharapkan dapat memberikan kesimpulan mengenai bahasa lokal di Indonesia yang memiliki tingkat kompressi yang baik untuk aplikasi potensial di komunikasi 5G.
- Selain itu, dengan teori informasi juga dapat diprediksi tingkat similaritas antar bahasa lokal, bahasa lokal dengan bahasa indonesia, Arab dan Inggris.
- Dalam konteks terorisme, komunikasi dengan "message" yang terkompress akan berdampak pada diprioritaskannya *end system* untuk mendapat kanal komunikasi, sementara yang tidak menggunakan kompressi khusus yang dilakukan orang yang *suspect* terrisme tidak akan mendapat layanan kanal komunikasi atau bahkan bisa diblok.
- Kompresi sendiri bertujuan agar pada proses transmisi data rekaman percakapan, data terenkripsi dan ukurannya menjadi lebih kecil sehingga berdampak pada *low power consumption.*
- Penelitian akan menjadi model integrasi teknik kompressi untuk berbagai bahasa lokal Indonesia, serta model dalam melihat tingkat kompressi yang optimal.
- Penelitian membantu penyelesaian masalah-masalah nasional terkait keamanan dan pencegahan terorisme.

BAB II. Tinjauan Pustaka

2.1. Tinjauan Literatur dan Riset Pendahuluan

Kompresi data adalah proses yang mengurangi ukuran data, menghilangkan informasi yang berlebihan. Ukuran data yang lebih pendek cocok karena menyiratkan pengurangan biaya. Tujuan kompresi data adalah untuk mengurangi redundansi pada data yang tersimpan atau dikomunikasikan, sehingga meningkatkan kepadatan data yang efektif. Kompresi data merupakan aplikasi penting di bidang penyimpanan file dan sistem terdistribusi karena pada data sistem terdistribusi harus dikirim dari dan ke seluruh sistem [5].

Di antara teknik kompresi, pengkodean Huffman adalah skema pengkodean awal yang optimal, yang menjamin *decodability* unik dari simbol yang terkompresi [3]. Kode tersebut dirumuskan oleh Huffman di awal tahun 1950an dan memberikan hasil yang optimal meski kodenya tidak universal [4]. Teknik lainnnya yang memberikan hasil lebih baik adalah Arithmetic encoding. Teknik ini memberikan rasio kompresi yang lebih baik dan ruang penyimpanan yang lebih rendah daripada Huffman Coding, walaupun dari kecepatan kompresi dan dekompressi masih lebih unggul Huffman coding [5] [6].

Algoritma lain adalah Run Length Encoding (RLE), yang merupakan teknik encoding sangat sederhana untuk data yang repetitif dan sequential [7].

Saat ini, berbagai teknik kompresi losless untuk data teks terus dikembangkan. Diantaranya adalah teknik yang dikembangkan oleh Ibrahim Akman yang mengimplementasikan sebuah algoritma kompresi teks lossless yang menggunakan morfologi bahasa multi-silabus berbasis suku kata. Algoritma yang diusulkan dirancang untuk mempartisi kata-kata menjadi suku kata dan kemudian menghasilkan representasi bit yang lebih pendek untuk kompresi [8]. Agus dalam [9] mengusulkan algoritma baru untuk kompresi data, yang disebut

jbit encoding (JBE). Algoritma ini akan memanipulasi setiap bit data dalam file untuk meminimalkan ukuran tanpa kehilangan data setelah decoding, oleh karenanya diklasifikasikan ke dalam kategori *lossless compression*.

Berbagai teknik kompressi data ini sudah dipakai secara luas dalam berbagai bidang dan jenis data, seperti data teks, audio, video, dan gambar. Diantara implementasinya adalah kompresi untuk data medis [10], kompresi suara untuk telekomunikasi dan kompresi data video untuk kebutuhan video streaming. Khusus untuk kompresi data teks, berbagai algoritma sudah diimplementasikan untuk melihat level kompresi berbagai bahasa. Berbagai pengembangan dan ujicoba tersebut dilakukan karena setiap bahasa punya keunikan simbol sendiri, dan berdampak pada *codeword* dan level kompresinya. Beberapa bahasa yang sudah analisis diantaranya adalah Bahasa Indonesia [11], Bengali [12] [13], Bahasa Sunda, Jawa, Bali, Inggris, dan Bahasa Prancis [3]. Pada penelitian [3], jumlah simbol yang tercover dalam naskah sumber masih terbatas, tidak semua ada.

Langkah awal dalam analisis naskah sumber untuk pengkodean Huffman digunakan analisis simbol yang ada pada naskah sumber dengan *marginal probability* untuk mengetahui frekuensi setiap simbol [14]. Menurut [14], jika diberikan *joint probability* kejadian-kejadian A dan B dari masing-masing variabel acak x dan y, sebagai $P(A, B)$, probabilitas kejadian A yang dihitung dalam bentuk,

$$P(A) = \sum_{y \in B} P(A, B) \quad (1)$$

disebut dengan *marginal probability*. Probabilitas ini akan menjadi dasar bagi perhitungan entropy masing-masing simbol dan entropy sumber, yang dinyatakan dalam bentuk,

$$H(X) = \sum_{x \in X} p(x) log\, p(x), \quad (2)$$

dimana X adalah variabel diskrit acak dan H(X) adalah entropy-nya.

[15]

Menurut [15], kesamaan bahasa akan berpengaruh pada efisiensi kompresi. Hal ini disebabkan karena dengan banyaknya kesamaan akan menekan jumlah *code word*, dan error yang terjadi lebih kecil sehingga komunikasi lebih efisien.

Untuk membandingkan level kompressi yang lebih baik, digunakan pendekatan *outage probability*. *Outage probability* merupakan perangkat penting untuk mengevaluasi kinerja komunikasi dalam *Rayleigh fading channels* [3]. *Outage probability* dari 2 bahasa dinyatakan dengan

$$P_{out} = A_1 + A_2, \tag{3}$$

dimana

$$A_1 = Pr[0 < R_X < H(X), 0 < R_Y], \tag{4}$$

dan

$$A_2 = Pr[H(B) < R_X, 0 < R_Y < H(Y)], \tag{5}$$

dimana R_X dan R_Y adalah laju kompresi untuk Bahasa X dan Bahasa Y.

Pada penelitian ini diusulkan untuk melakukan analisis tingkat kompresi berbagai bahasa lokal Indonesia (Sunda, Jawa, Bali, Minang, Bugis), Inggris, dan Arab Pegon untuk melihat manakah yang memiliki tingkat kompresi lebih baik. Penelitian akan didasarkan pada naskah dengan ukuran relatif besar, agar semua simbol di setiap bahasa dapat dicover. Penelitian juga akan meningatkan level kompresi dengan *Run Length Encoding* (RLE) jika persyaratan awal dengan melihat nilai entropy sumber berdasarkan Huffman Coding lebih kecil dari 5.

2.2. Road Map Penelitian

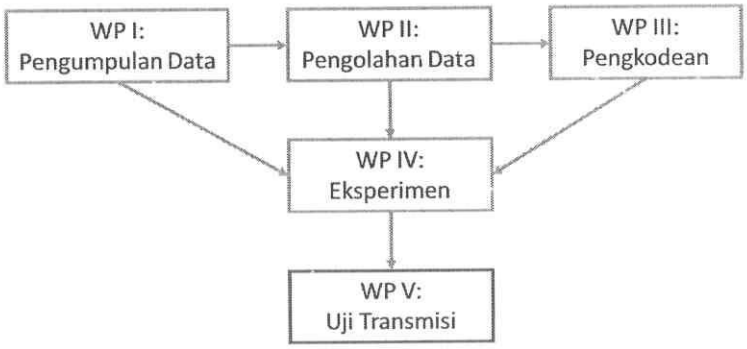
Berikut ini adalah peta jalan penelitian. Penelitian awal sedang berjalan di tahun 2017, dengan melakukan analisis terhadap bahasa.

Gambar 2. Roadmap penelitian

BAB III. Metodologi Penelitian

3.1. Pendekatan dan Pentahapan

Penelitian ini merupakan kegiatan analisis dan eksperimental. Ada 3 tahapan utama dalam penelitian ini, sebagaimana diperlihatkan oleh gambar 3.



Gambar 3. Tahapan pekerjaan

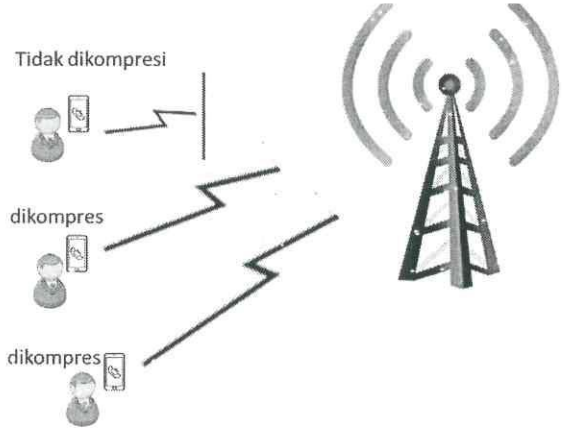
Working Project I (Pengumpulan data): Tahapan ini merupakan tahap awal penelitian, yaitu pengumpulan data-data yang diperlukan. Data diperoleh dengan mengumulkan naskah berbahasa daerah. Jumlah bahasa daerah yang digunakan lebih banyak dari yang pernah dilakukan oleh Khaoirul Anwar pada [1]. Bahasa daerah yang dianalisis adalah Bahasa Sunda, Jawa, dengan pertimbangan bahasa tersebut adalah bahasa dengan suku dominan di Indonesia.
Working Project II (Pengolahan): Teori informasi banyak dipakai dalam melakukan analisis terhadap data dan sinyal. Sebelum dilakukan analisis dengan teori informasi, terlebih dahulu harus dipastikan dan didefinisikan mengenai berbagai simbol/karakter bahasa lokal yang ada, bahasa inggris, dan arab pegon. Selanjutnya, pada tahap ini, setiap simbol akan diolah berdasarkan teori informasi, mulai dari penghitungan frekuensi simbol (*marginal probability*), sampai nilai entropy sumber. Perhitungan awal akan menggunakan pendekatan

marginal probability untuk melihat frekuensi kemunculan simbol. Pada tahapan ini juga dilakukan analisis kemiripan bahasa menggunakan *joint probability*.

Working Project III (Pengkodean): Salahsatu implementasi teori informasi diterapkan pada teknik kompressi. Huffman code sampai saat ini merupakan salahsatu teknik kompressi yang optimal. Berbagai bahasa lokal, bahasa indonesia, bahasa arab, dan bahasa inggris akan dianalisis dengan *binary huffman code* berdasarkan data probabilitas, dan entropy yang sudah diolah. Sebagai pembanding, analisis juga dilakukan dengan kombinasi *run length-huffman code* untuk meningkatkan level kompresi dan *ternary huffman code*. Berbagai variasi teknik kompressi dan variasi bahasa ini akan menemukan tingkat kompressi yang optimal itu untuk berbagai bahasa dan berbagai teknik pengkodean. Untuk membandingkan tingkat kompresi yang lebih optimal diantara berbagai bahasa yang dianalisis, digunakan pendekatan *outage probability*.

Working Project IV (Eksperimen): Pada tahap ini akan dilakukan pengujian dengan data text dari naskah yang berbeda berdasarkan pola yang sudah ditemukan pada tahapan sebelumnya. Pengujian dilakukan menggunakan simulasi software sampai dihasilkan prototipe Alpha (skala lab) dari sistem yang dikembangkan.

Working Project V (Uji transmisi): Tahapan ini dilakukan pada **tahun kedua**, dengan melakukan pengujian pada sistem transmisi sampai diperoleh prototipe Betha dari sistem.



Gambar 4. Skema penelitian

3.2. Data dan Sumber Data

Data awal penelitian berupa tulisan/naskah dengan berbagai bahasa lokal (Sunda latin, Sunda aksara, Jawa latin, Jawa aksara) dengan jumlah simbol memenuhi kemunculan semua simbol.

BAB IV. Analisis Statistik Berbagai Bahasa

4.1. Bahasa Jawa Beraksara Jawa

Akasara Jawa merupakan bagian Bahasa Jawa yang teresap dari tulisan sansakerta dan palapa. Aksara Jawa memiliki banyak karakter dan simbol. Aksara jawa mempunyai bentuk kombinasi dan penulisan yang unik dimana dapat disisipkan dan ditumpukan. Aksara jawa dalam sturktur bahasanya terbagi atas beberapa bagian, seperti aksara jawa carakan (ngalegena), aksara jawa pasangan (mati), aksara suaara, aksara rekan, aksara murda, angka/wilangan jawa, tanda baca (sandangan), juga akasara jawa yang terkombinasi. Aksara Jawa dalam penulisan huruf mempunyai dua tipe, yaitu aksara Jawa dengan huruf besar yang kedua huruf aksara Jawa kecil.

Cacarakan atau carakan merupakan aksara Jawa modern hasil modifikasi dari aksara kawi. Hal ini dapat dilihat dari struktur huruf yang tidak dikatakan 2 buah huruf aksara dalam huruf latin jawa. Contoh aksara *RA* jika dilatinkan maka memiliki 2 huruf yaitu R dan A [13]. Aksara Jawa memiliki 20 karakter dasar (aksara jawa carakan). Contoh Karakter-karakter Aksara jawa carakan dapat dilihat pada Tabel 1.

Tabel 1. Akasara Jawa Carakan

Aksara Jawa Carakan (ngalegena)								
No.	Huruf	Aksara Jawa	No.	Huruf	Aksara Jawa	No.	Huruf	Aksara Jawa
1.	ha	ꦲ	8.	da	ꦢ	15.	pa	ꦥ
2.	na	ꦤ	9.	ta	ꦛ	16.	dha	ꦝ
3.	ca	ꦕ	10.	sa	ꦱ	17.	ja	ꦗ
4.	ra	ꦫ	11.	wa	ꦮ	18.	ya	ꦪ

Aksara Jawa Carakan (ngalegena)								
No.	Huruf	Aksara Jawa	No.	Huruf	Aksara Jawa	No.	Huruf	Aksara Jawa
5.	ka		12.	la		19.	nya	
6.	ma		13.	ga		20.	ba	
7.	tha		14.	nga				

Aksara jawa dalam karakternya memiliki huruf besar dan huruf kecil. Contoh karakter aksara jawa carakan besar dapat dilihat pada Tabel 2.

Tabel 2. Aksara Jawa Carakan besar

Aksara Jawa Carakan (ngalegena) besar								
No.	Huruf	Aksara Jawa	No.	Huruf	Aksara Jawa	No.	Huruf	Aksara Jawa
1.	Ha		8.	Da		15.	Pa	
2.	Na		9.	Ta		16.	Dha	
3.	Ca		10.	Sa		17.	Ja	
4.	Ra		11.	Wa		18.	Ya	
5	Ka		12.	La		19.	Nya	
6.	Ma		13.	Ga		20.	Ba	
7.	Tha		14.	Nga				

Aksara jawa memiliki 20 karakter pasangan. Contoh karakter-karakter aksara jawa pasangan (mati) dapat dilihat pada Tabel 3.

Tabel 3. Aksara Jawa Pasangan (mati)

Aksara Jawa Pasangan (mati)								
No.	Huruf	Aksara Jawa	No.	Huruf	Aksara Jawa	No.	Huruf	Aksara Jawa
1.	h		8.	d		15.	p	
2.	n		9.	t		16.	dh	
3.	c		10.	s		17.	j	
4.	r		11.	w		18.	y	
5.	k		12.	l		19.	ny	
6.	m		13.	g		20.	b	
7.	th		14.	ng				

Aksara jawa pasangan digunakan pada saat penulisan dengan akhiran kata mati. Kata mati maksudnya merupakan suatu kata dari kalimat atau *paragraph* yang diakhiri bukan dengan karakter swara atau *vocal.*

Aksara jawa memiliki 5 karakter swara dasar. Contoh karakter-karakter aksara jawa suara dapat dilihat pada Tabel 4.

Tabel 4. Aksara Jawa Swara

Aksara Jawa Swara								
No.	Huruf	Aksara Jawa	No.	Huruf	Aksara Jawa	No.	Huruf	Aksara Jawa
1.	A		3.	I		5.	U	
2.	E		4.	O				

Aksara jawa memiliki beberapa karakter sandhangan. Contoh Karakter-karakter aksara Jawa sandangan dapat dilihat pada Tabel 5.

Tabel 5. Aksara Jawa Sandangan

Tanda Baca Sandangan								
No.	Huruf	Aksara Jawa	No.	Huruf	Aksara Jawa	No.	Huruf	Aksara Jawa
1.	a		12.	I	ꦶ	23.	u	ꦸ
2.	e	ꦺ	13.	O	ꦺꦴ	24.	ė	ꦺꦀ
3.	_r	ꦽ	14.	_h	ꦲ꧀	25.	_ng	ꦁ
4.	_ya	ꦾ	15.	_ra	ꦿ	26.	_re	ꦿꦽ
5.	_		16.	:	꧈	27.	‘	‘
6.	.	꧉	17.	(	꧁	28.	/	꧇
7.	,	꧈	18.	)	꧂	29.	Pada luhur	ꦄ
8.	adeg-adeg	꧋	19.	Wasana pada	ꦄ ꦮ ꦄ	30.	Pada andap	ꦄ
9.	Pada pangkat	꧋ꦴ꧋	20.	Purwa pada	ꦄ ꦄꦱ꦳ꦄ			
10.	Pada guru	꧉ꦴ꧉	21.	Madya pada	ꦄ꦳ ꦄ			
11.	Pada madya	ꦄ꦳	22.	Pada pancake	꧉ꦴ꧉			

Aksara jawa memiliki 10 karakter wilangan. Contoh karakter-karakter aksara jawa wilangan dapat dillihat pada Tabel 6.

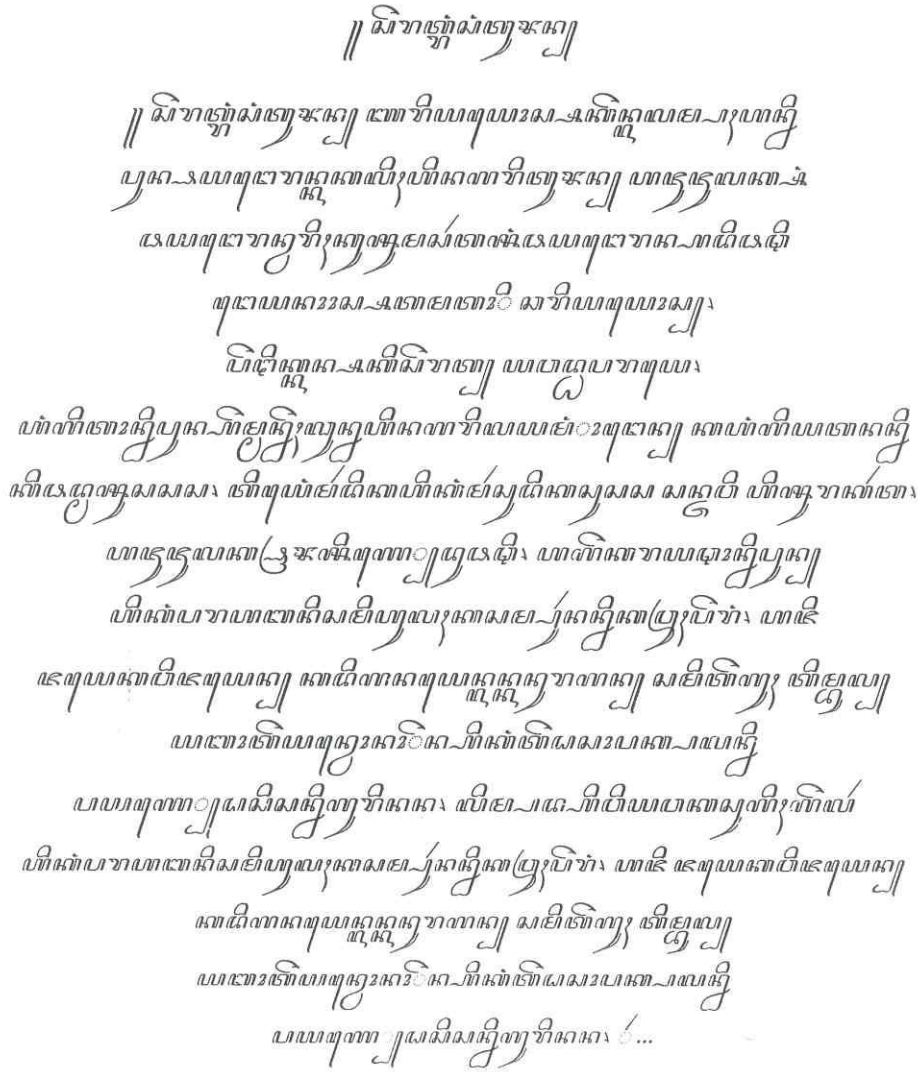
Tabel 6. Aksara Wilangan

Aksara Jawa wilangan								
No.	Huruf	Aksara Jawa	No.	Huruf	Aksara Jawa	No.	Huruf	Aksara Jawa
1.	1	꧑	5.	5	꧕	9.	9	꧙
2.	2	꧒	6.	6	꧖	10.	0	꧐
3.	3	꧓	7.	7	꧗			
4.	4	꧔	8.	8	꧘			

Aksara Jawa diakumulasikan mempunyai 20 karakter dasar (Aksara Carakan), 20 huruf pasangan yang digunakan untuk menutup huruf swara. 5 karakter suara (vocal depan), 5 huruf rekan dan pasangannya, 10 karakter aksara Jawa wilangan. Aksara Jawa didalam penulisannya memilliki karakter dengan huruf besar dan kecil. Aksara jawa memiliki penulisan yang unik, dimana dapat disisipkan, ditambahi dan ditumpukan. Selain huruf dan angka, aksara jawa memiliki banyak tanda baca. Aksara jawa juga mempunyai beberapa huruf sandhagan, beberapa karakter khusus, dan beberapa karakter yang digunakan sebagai pengatur penulisan.

4.1.1 Objek Yang Dianalisis

Objek yang digunakan sebagai sampel adalah teks aksara jawa. Teks aksara jawa diperoleh dari sebuah artikel blog. Judul cerita yang digunakan sebagai sampel adalah 'Serat Rangsangan Tuban','Babading Kabudayan Jawi','Pengetan', dan lain-lain. Sampel penelitian diambil dari web blog yang dimuat pada tahun 2012 oleh R. S. Wihanato. Kutipan artikel teks Bahasa Jawa beraksara yang digunakan,



Sumber: https://sites.google.com/site/jawaunicode/download

4.1.2 Pengumpulan dan Pengolahan Data

Data yang dikumpulkan dari teks aksara jawa diakumulasikan pada Tabel 7, dimana diasumsikan simbol-simbol (*S*), frekuensi (*F(x)*), probabilitas kemunculan (*p(x)*), *entropy* (*H(x)*), *x* merupakan *random varieble*.

Tabel 7. Pengumpulan dan Pengolahan data Bahasa Jawa Beraksara Jawa

S	*F(x)*	*sorting*		*H(x)*
		S	*p(x)*	
ha	51	spasi	0.1680	0.4323
na	272	i	0.0806	0.2928
ca	42	u	0.0572	0.2361
ra	249	ka	0.0454	0.2025
ka	404	n	0.0409	0.1886
da	170	ng	0.0354	0.1706
ta	275	e	0.0325	0.1607
sa	255	ta	0.0309	0.1550
wa	209	na	0.0305	0.1536
la	164	é	0.0296	0.1503
pa	240	sa	0.0286	0.1467
dha	2	ra	0.0280	0.1444
ja	99	pa	0.0269	0.1403
ya	113	ma	0.0248	0.1323
nya	6	wa	0.0235	0.1272
ma	221	da	0.0191	0.1091
ga	133	,	0.0185	0.1065
ba	130	la	0.0184	0.1061
tha	1	a	0.0183	0.1056
nga	70	o	0.0163	0.0968
h	79	_ng	0.0162	0.0964
n	364	ga	0.0149	0.0904
c	1	ba	0.0146	0.0890
r	20	ya	0.0127	0.0800
k	39	ja	0.0111	0.0721
d	22	h	0.0089	0.0606
t	45	s	0.0085	0.0585
s	76	_ra	0.0085	0.0585
w	1	_r	0.0083	0.0574

S	*F(x)*	*sorting*		*H(x)*
		S	*p(x)*	
l	20	nga	0.0079	0.0552
p	12	.	0.0070	0.0501
dh	1	Pa	0.0068	0.0490
j	1	Sa	0.0067	0.0484
y	1	m	0.0058	0.0431
ny	1	ha	0.0057	0.0425
m	52	t	0.0051	0.0388
g	6	Ra	0.0051	0.0388
b	9	ca	0.0047	0.0363
th	1	k	0.0044	0.0344
ng	315	Ka	0.0044	0.0344
a	163	adeg-adeg	0.0035	0.0286
i	718	Ta	0.0033	0.0272
u	509	Ma	0.0030	0.0251
e	289	Wa	0.0027	0.0230
o	145	d	0.0025	0.0216
é	264	Da	0.0025	0.0216
Ha	12	r	0.0022	0.0194
Na	12	I	0.0022	0.0194
Ca	4	Ja	0.0022	0.0194
Ra	45	A	0.0022	0.0194
Ka	39	pada pangkat	0.0020	0.0179
Da	22	Nga	0.0018	0.0164
Ta	29	1	0.0018	0.0164
Sa	60	Ba	0.0017	0.0156
Wa	24	p	0.0013	0.0125
La	1	Ha	0.0013	0.0125
Pa	61	Na	0.0013	0.0125
Dha	1	I	0.0012	0.0116
Ja	20	:	0.0012	0.0116
Ya	1	2	0.0011	0.0108
Nya	1	b	0.0010	0.0100
Ma	27	8	0.0009	0.0091
Ga	4	3	0.0008	0.0082
Ba	15	Strip (-)	0.0008	0.0082
Tha	1	nya	0.0007	0.0073

S	*F(x)*	*sorting*		*H(x)*
		S	*p(x)*	
Nga	16	g	0.0007	0.0073
A	20	_ya	0.0007	0.0073
I	11	)	0.0007	0.0073
U	2	pada luhur/madya	0.0007	0.0073
E	2	1a	0.0007	0.0073
O	1	Ca	0.0004	0.0045
Ė	4	Ga	0.0004	0.0045
_r	74	Ė	0.0004	0.0045
_h	1	_re	0.0004	0.0045
_ng	144	4	0.0004	0.0045
_ya	6	5	0.0004	0.0045
_ra	76	0	0.0004	0.0045
_re	4	(	0.0004	0.0045
1	16	6	0.0003	0.0035
2	10	7	0.0003	0.0035
3	7	9	0.0003	0.0035
4	4	3a	0.0003	0.0035
5	4	0a	0.0003	0.0035
6	3	I	0.0003	0.0035
7	3	dha	0.0002	0.0025
8	8	U	0.0002	0.0025
9	3	E	0.0002	0.0025
0	4	(*)	0.0002	0.0025
Strip (-)	7	2a	0.0002	0.0025
.	62	8a	0.0002	0.0025
,	165	9a	0.0002	0.0025
:	11	tha	0.0001	0.0013
(	4	c	0.0001	0.0013
)	6	w	0.0001	0.0013
(')	1	dh	0.0001	0.0013
/	1	j	0.0001	0.0013
(*)	2	y	0.0001	0.0013
adeg-adeg	31	ny	0.0001	0.0013
pada pangkat	18	th	0.0001	0.0013
pada guru	1	La	0.0001	0.0013
pada pancak	1	Dha	0.0001	0.0013

S	*F(x)*	*sorting*		*H(x)*
		S	*p(x)*	
pada luhur/madya	6	Ya	0.0001	0.0013
purwa pada	1	Nya	0.0001	0.0013
madya pada	1	Tha	0.0001	0.0013
wasana pada	1	O	0.0001	0.0013
1a	6	_h	0.0001	0.0013
2a	2	(')	0.0001	0.0013
3a	3	/	0.0001	0.0013
4a	1	pada guru	0.0001	0.0013
5a	1	pada pancak	0.0001	0.0013
6a	1	purwa pada	0.0001	0.0013
7a	1	madya pada	0.0001	0.0013
8a	2	wasana pada	0.0001	0.0013
9a	2	4a	0.0001	0.0013
0a	3	5a	0.0001	0.0013
V	1	6a	0.0001	0.0013
I	3	7a	0.0001	0.0013
N	1	V	0.0001	0.0013
V (romawi)	1	N	0.0001	0.0013
R	1	V	0.0001	0.0013
spasi	1496	R	0.0001	0.0013
Jumlah				
120	8906		1	5.0480

Hasil pengumpulan data untuk simbol yang digunakan aksara Jawa (*S*) 120 simbol, banyak data (N) 8906 karakter. Artikel yang digunakan tidak hanya merangkum karakter aksara Jawa tetapi merangkum karakter yang bukan aksara Jawa. Karakter-karakter yang bukan aksara jawa ini melambangkanseperti karakter-karakter bahasa Indonesia. Karakter-karakter yang ada pada artikel yang bukan aksara jawa 1a, 2a, 3a, 4a, 5a, 6a, 7a, 8a, 9a, 0a, V, I, N, V (romawi), R. Inisialisasi dilakukan pada saat pengumpulan data. Inisisalisasi terjadi karna artikel yang digunakan tidak dapat merangkum semua karakter aksara jawa. Inisialisasi dilakukan dengan cara pemberian frekuensi pada karakter-karakter

BAB V. Analisis Tingkat Kompresi dan Kinerja Bahasa

5.1. Menghitung Expected Code Length

Expected c*ode length* didapatkan dari hasil perhitungan dengan menggunakan Persamaan 6.

$$L(C(a)) = \sum_{x \in X} l(x)p(x) \quad (6)$$

Expected code length merupakan jumlah hasil dari perkalian antara panjang codewords *(l(x)* dengan *entropy H(x)* dari tiap karakter yang diberikan satuan bits. Nilai *expected code length* berperan dalam penentuan efisien. *Expected code length* dikatakan baik ketika nilai *Expected code length* mendekati *entropy.* Semakin kecil nilai perbandingan *Expected code length* dengan *entropy* maka semakin besar efisiensi yang didapat.

5.1.1. Bahasa Jawa Aksara Jawa

Data *expected code length* yang didapatkan dari teks aksara Jawa diakumulasikan pada Tabel 23, dimana diasumsikan simbol-simbol *(S)*, probabilitas kemunculan *(p(x))*, dan *entropy (H(x))*, *codewords (C(x)), Panjang codewords (l(x)), expected code Length (L(C)), x* merupakan *random variable.*

Tabel 23. *Expected Code Length* Aksara Jawa

Sorting		*H(X)*	*C(x)*	*l(x)*	*L(C)*
S	*p(x)*				
Spasi	0.1680	0.4323	100	3	0.5040
I	0.0806	0.2928	010	3	0.2418
U	0.0572	0.2361	0101	4	0.2288
Ka	0.0454	0.2025	10110	5	0.2270
N	0.0409	0.1886	0110	4	0.1636
Ng	0.0354	0.1706	11010	5	0.1770
E	0.0325	0.1607	11011	5	0.1625
Ta	0.0309	0.1550	111110	6	0.1854
Na	0.0305	0.1536	111111	6	0.1830

É	0.0296	0.1503	01000	5	0.1480
Sa	0.0286	0.1467	00100	5	0.1430
Ra	0.0280	0.1444	00101	5	0.1400
Pa	0.0269	0.1403	00110	5	0.1345
Ma	0.0248	0.1323	101001	6	0.1488
Wa	0.0235	0.1272	101010	6	0.1410
Da	0.0191	0.1091	101011	6	0.1146
,	0.0185	0.1065	01111	5	0.0925
la	0.0184	0.1061	110000	6	0.1104
a	0.0183	0.1056	110001	6	0.1098
o	0.0163	0.0968	111000	6	0.0978
_ng	0.0162	0.0964	111001	6	0.0972
ga	0.0149	0.0904	000010	6	0.0894
ba	0.0146	0.0890	000011	6	0.0876
ya	0.0127	0.0800	100000	6	0.0762
ja	0.0111	0.0721	1011101	7	0.0777
h	0.0089	0.0606	1100100	7	0.0623
s	0.0085	0.0585	1100101	7	0.0595
_ra	0.0085	0.0585	1100110	7	0.0595
_r	0.0083	0.0574	1110100	7	0.0581
nga	0.0079	0.0552	1110101	7	0.0553
.	0.0070	0.0501	0011100	7	0.0490
Pa	0.0068	0.0490	0011101	7	0.0476
Sa	0.0067	0.0484	10100010	8	0.0536
m	0.0058	0.0431	10111000	8	0.0464
ha	0.0057	0.0425	10111001	8	0.0456
t	0.0051	0.0388	10111100	8	0.0408
Ra	0.0051	0.0388	10111101	8	0.0408
ca	0.0047	0.0363	0111010	7	0.0329
k	0.0044	0.0344	0111010	7	0.0308
Ka	0.0044	0.0344	11101100	8	0.0352
adeg-adeg	0.0035	0.0286	11101101	8	0.0280
Ta	0.0033	0.0272	00111100	8	0.0264
Ma	0.0030	0.0251	101000110	9	0.0270
Wa	0.0027	0.0230	101101100	9	0.0243
d	0.0025	0.0216	101101101	9	0.0225
Da	0.0025	0.0216	01110010	8	0.0200

r	0.0022	0.0194	01110011	8	0.0176
l	0.0022	0.0194	110011100	9	0.0198
Ja	0.0022	0.0194	110011101	9	0.0198
A	0.0022	0.0194	111011100	9	0.0198
pada pangkat	0.0020	0.0179	111011101	9	0.0180
Nga	0.0018	0.0164	111011110	9	0.0162
l	0.0018	0.0164	111011111	9	0.0162
Ba	0.0017	0.0156	001111010	9	0.0153
p	0.0013	0.0125	1011111100	10	0.0130
Ha	0.0013	0.0125	1011111101	10	0.0130
Na	0.0013	0.0125	011100000	9	0.0117
I	0.0012	0.0116	011100001	9	0.0108
:	0.0012	0.0116	011100010	9	0.0108
2	0.0011	0.0108	011100011	9	0.0099
b	0.0010	0.0100	10111111100	11	0.0110
8	0.0009	0.0091	10111111101	11	0.0099
3	0.0008	0.0082	0011110110	10	0.0080
Strip (-)	0.0008	0.0082	0011110111	10	0.0080
nya	0.0007	0.0073	10100011100	11	0.0077
g	0.0007	0.0073	10100011101	11	0.0077
_ya	0.0007	0.0073	10100011110	11	0.0077
)	0.0007	0.0073	10100011111	11	0.0077
pada uhur/andhap	0.0007	0.0073	10110111100	11	0.0077
la	0.0007	0.0073	10110111101	11	0.0077
Ca	0.0004	0.0045	00111110000	11	0.0044
Ga	0.0004	0.0045	00111110001	11	0.0044
É	0.0004	0.0045	00111110010	11	0.0044
_re	0.0004	0.0045	00111110011	11	0.0044
4	0.0004	0.0045	00111110100	11	0.0044
5	0.0004	0.0045	00111110101	11	0.0044
0	0.0004	0.0045	00111110110	11	0.0044
(	0.0004	0.0045	00111110111	11	0.0044
6	0.0003	0.0035	101111111100	12	0.0036
7	0.0003	0.0035	101111111101	12	0.0036
9	0.0003	0.0035	101111111110	12	0.0036
3a	0.0003	0.0035	101111111111	12	0.0036

0a	0.0003	0.0035	110011111000	12	0.0036
I	0.0003	0.0035	110011111001	12	0.0036
dha	0.0002	0.0025	110011111010	12	0.0024
U	0.0002	0.0025	110011111011	12	0.0024
E	0.0002	0.0025	110011111100	12	0.0024
(*)	0.0002	0.0025	110011111101	12	0.0024
2a	0.0002	0.0025	110011111110	12	0.0024
8a	0.0002	0.0025	110011111111	12	0.0024
9a	0.0002	0.0025	001111110000	12	0.0024
tha	0.0001	0.0013	0011111100010	13	0.0013
c	0.0001	0.0013	0011111100011	13	0.0013
w	0.0001	0.0013	0011111100100	13	0.0013
dh	0.0001	0.0013	0011111100101	13	0.0013
j	0.0001	0.0013	0011111100110	13	0.0013
y	0.0001	0.0013	0011111100111	13	0.0013
ny	0.0001	0.0013	0011111101000	13	0.0013
th	0.0001	0.0013	0011111101001	13	0.0013
La	0.0001	0.0013	0011111101010	13	0.0013
Dha	0.0001	0.0013	0011111101011	13	0.0013
Ya	0.0001	0.0013	0011111101100	13	0.0013
Nya	0.0001	0.0013	0011111101101	13	0.0013
Tha	0.0001	0.0013	0011111101110	13	0.0013
O	0.0001	0.0013	0011111101111	13	0.0013
_h	0.0001	0.0013	0011111110000	13	0.0013
(')	0.0001	0.0013	0011111110001	13	0.0013
/	0.0001	0.0013	0011111110010	13	0.0013
pada guru	0.0001	0.0013	0011111110011	13	0.0013
pada pancak	0.0001	0.0013	0011111110100	13	0.0013
purwa pada	0.0001	0.0013	0011111110101	13	0.0013
madya pada	0.0001	0.0013	0011111110110	13	0.0013
wasana pada	0.0001	0.0013	0011111110111	13	0.0013
4a	0.0001	0.0013	0011111111000	13	0.0013
5a	0.0001	0.0013	0011111111001	13	0.0013
6a	0.0001	0.0013	0011111111010	13	0.0013
7a	0.0001	0.0013	0011111111011	13	0.0013

V	0.0001	0.0013	0011111111100	13	0.0013
N	0.0001	0.0013	0011111111101	13	0.0013
V	0.0001	0.0013	0011111111110	13	0.0013
R	0.0001	0.0013	0011111111111	13	0.0013
	1	5.0480			5.1548

Expected code length untuk Bahasa Jawa beraksara jawa *L(C(a))* adalah 5.1548 bits. Panjang *codewords* yang paling terpanjang pada *codeword* Bahasa Jawa beraksara jawa sebanyak 13 (*constraint*). *Expected code length* dihitung dengan menggunakan Persamaan 6. Perhitungan *Expected code length* Bahasa Jawa beraksara jawa *L(C(a))* adalah sebagai berikut:

$$L(C(a)) = \sum_{x \in X} l(x)p(x) = (3 * 0.1680) + (3 * 0.806) + (4 * 0.0572) + \cdots + (0.0001 * 13)$$

$$= 5.1548 \text{ bits}$$

5.1.2. Bahasa Jawa Latin Jawa

Data *expected code length* yang didapatkan dari teks Jawa latin diakumulasikan pada Tabel 24, dimana diasumsikan simbol-simbol (*S*), probabilitas kemunculan (*p(x)*), dan *entropy* (*H(X)*), *Codewords* (*C(x)*), Panjang *codewords* (*l(x)*), *expected code Length (L(C))*, *x* merupakan *random variable*.

Tabel 24. *Expected Code Length* Latin Jawa

Sorting		***H(X)***	***C(x)***	***l(x)***	***L(C)***
S	***p(x)***				
Spasi	0.1427	0.3899	001	3	0.4281
A	0.1350	0.0827	011	3	0.4050
N	0.0939	0.0439	111	3	0.2817
I	0.0700	0.1410	0101	4	0.2801
E	0.0640	0.2539	1010	4	0.2560
G	0.0513	0.0237	1001	4	0.2052

BAB VI. Kesimpulan dan Saran

6.1. Kesimpulan

Berdasarkan penelitian yang telah dilakukan dapat disimpulkan sebagai berikut:

1. Penelitian ini memprediksikan bahwa Bahasa Sunda mempunyai nilai *Entropy* dan *Compression Rates* jauh lebih rendah dibandingkan dengan Aksara Sunda latin, dimana nilai *Entropy* dan *Compression Rates* Bahasa Sunda bernilai sebesar 4.58 bit/karakter dan 0.5775 sedangkan Aksara Sunda bernilai 6.00 bit/karakter dan 0.6698.
2. Kinerja Pengkompresian dengan menggunakan *Huffman Codes* pada bahasa daerah berkerja dengan baik dan efisien, hal itu dapat dilihat dari hasil kompresi. Dimana Aksara Sunda dikompres sebesar 6 bit/karakter dari kode asli (Expected Code Length Standar) yaitu sebesar 9 bit/karakter dan Bahasa Sunda dapat dikompres sebesar 4 bit/karakter dari kode asli (ASCII Codes) yaitu sebesar 8 bit/karakter.
3. Bahasa Jawa latin jawa mempunyai nilai *entropy* dan *compression rates* lebih kecil dibanding dengan Bahasa Jawa aksara jawa. Bahasa jawa latin jawa mempunyai *entropy* 4.4118 bits dan *compression rates* 0.6325, sedangkan Bahasa Jawa aksara jawa *entropy* 5.0480 bits dan *compression rates* 0.7364.
4. Bahasa Jawa latin jawa mendapatkan efisiensi lebih besar dibanding dengan Bahasa Jawa aksara jawa. Bahasa Jawa latin Jawa mempunyai efisiensi 99.65% sedangkan Bahasa Jawa mempunyai efisiensi 97.93 %
5. Setelah dilakukan pengujian dengan uji teoritis menggunakan *Outage Probability*, tingkat kinerja komunikasi Aksara Sunda dan Bahasa Sunda tidak akan terdapat *error floor* jika kedua *user* berada dalam kondisi bergerak secara bersamaan. Namun, jika salah satu *user* beada dalam kondisi diam, maka akan terdapat *error floor*.

6.2. Saran

Untuk pengembangan riset selanjtutnya disarankan:

1. Menggunakan sampel dengan jumlah simbol lebih banyak
2. Melakukan analisis terhadap bahasa daerah lain yang potensial
3. Melakukan simulasi dengan user yang bervariasi

REFERENSI

[1] D. Salomon, Data Compression: The Complete Reference, 4th Edition, New York, USA: Springer, 2011, pp. 2-15; 51-124.

[2] E. Capo-chichi, H. Guyennet and J. Friedt, "RLE -A New Data Compression Algorithm for Wireless Sensor Network," in *Third International Conference on Sensor Technologies and*, 2009.

[3] K. Anwar, R. F. Baihaqi and Y. Julian, "Source Coding-based Compressions of Indonesian Local Languages for 5G Potential Applications," in *International Symposium on Electronics and Smart Devices (ISESD)*, Yogyakarta, Indonesia, 2017.

[4] D. A. Huffman, "A Method for the Construction of Minimum-Redundancy Codes," in *Proceedings of IRE*, Cambridge, 1952.

[5] S. Porwal, Y. Chaudhary, J. Joshi and M. Jain, "Data Compression Methodologies for Lossless Data and Comparison between Algorithms," *International Journal of Engineering Science and Innovative Technology (IJESIT),* vol. 2, no. 2, pp. 142-147, 2013.

[6] J. Wang, X. Ji, S. Zhao, X. Xie and J. Kuang, "Context-based adaptive arithmetic coding in time and frequency domain for the lossless compression of audio coding parameters at variable rate," *EURASIP Journal on Audio, Speech, and Music Processing,* vol. 1, no. 9, pp. 1-13, 2013.

[7] S. Rao and P. Bhat, "Evaluation of Lossless Compression Techniques," in *IEEE ICCSP*, Melmaruvathur, 2015.

[8] I. Akman, H. Bayindir, S. Ozleme, Z. Akin and a. S. Misra, "Lossless Text Compression Technique Using Syllable Based Morphology," *The International Arab Journal of Information Technology,* vol. 8, no. 1, pp. 66-74, 2011.

[9] I. M. A. D. Suarjaya, "A New Algorithm for Data Compression Optimization," *International Journal of Advanced Computer Science and Applications (IJACSA),* vol. 3, no. 8, pp. 14-17, 2012.

[10] T. Dutta, "Medical Data Compression and Transmission in Wireless Ad Hoc Networks," *IEEE SENSORS JOURNAL,* vol. 15, no. 2, pp. 778-786, 2015.

[11] A. Sinaga, Adiwijaya and H. Nugroho, "Development of Word-Based Text Compression Algorithm for Indonesian Language Document," in *3rd International Conference on Information and Communication Technology (ICoICT)*, Denpasar, 2015.

[12] M. R. Islam and S. A. A. Rajon, "On the Design of an Effective Corpus for Evaluation of Bengali Text Compression Schemes," in *11th International Conference on Computer and Information Technology (ICCIT)*, Khulna, 2008.

[13] A. S. M. Arif, A. Mahamud and R. Islam, "An Enhanced STatic Data COmpression Scheme of Bengali Short Message," *International Journal of Computer Science and Information Security (IJCSIS),* vol. 4, no. 1, pp. 97-103, 2009.

[14] T. M. Cover and J. A. Thomas, Elements of Information Theory, New Jersey: John Wiley & Sons, 2006.

[15] C. Cui, Z. Dang, T. R. Fischer and O. H. Ibarra, "Similarity in languages and programs," *Theoretical Computer Science,* vol. 498, pp. 58-75, 2013.

Indeks

Kompresi Berbagai Bahasa Lokal Indonesia Berbasis Teori Informasi & Coding

Dr. Abdul Kodir
Nanang Ismail, MT
Asep Solih Awaluddin, M.Si
Prof. Dr. Uus Ruswandi, M.Pd

PUSAT PENELITIAN DAN PENERBITAN
UIN SGD BANDUNG